Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 554−561
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1973

Analysis of Community Sentiment on Twitter towards COVID-19 Vaccine Booster Using Ensemble Bagging Methods

Artamira Rizqy Amartya Maden, Jondri*, Widi Astuti

School of Computing, Informatics Study Program, Telkom University, Bandung, Indonesia
Email: [1]artamira@student.telkomuniversity.ac.id, [2,*]jondri@telkomuniversity.ac.id, [3]astutiwidi@telkomuniversity.ac.id
Email Penulis Korespondensi: jondri@telkomuniversity.ac.id
Submitted:**26/07/2022**; Accepted:**18/08/2022**; Published**: 30/09/2022**

Abstract− COVID-19 is an infectious disease caused by a newly discovered type of coronavirus. Based on recommendations from the Technical Advisory Group on Virus Evolution, WHO established a new variant called Omicron. Due to the rapid spread of COVID-19, a booster vaccine was created to deal with the new virus variant. However, the strategy of giving vaccines that never ends is considered controversial by the community, and this is shown by the number of people who express their opinions, both positive and negative opinions on social media, one of which is Twitter. This research was conducted by collecting data with the help of the Twitter API. Sentiment analysis is used to determine the distribution of positive and negative sentiments. After knowing the distribution of tweets, classification is carried out using the ensemble bagging method to determine the method's performance. The classification method uses ensemble bagging with three basic lessons, namely Naive Bayes, K-Nearest Neighbor, and Decision Tree. Meanwhile, the feature extraction used in this research is TF-IDF (Term Frequency-Inverse Document Frequency). The performance of the ensemble bagging method by applying Hyperparameter Tuning is a precision of 0.72, recall of 0.71, F1-Score of 0.72, and accuracy of 0.72.

Keywords: Vaccine; Booster; Sentiment; Baging; Twitter

1. INTRODUCTION

COVID-19 is an infectious disease caused by a newly discovered type of coronavirus. The epicentre of the spread of this virus began in Wuhan, China, in December 2019 [1]. President Joko Widodo announced that there were 2 Indonesian citizens who were confirmed positive for COVID-19 on March 2, 2020. As of February 16, 2022, the Indonesian government reported 4,966,046 confirmed cases of COVID-19 from 510 districts in 34 provinces [2]. The spread of COVID-19 is very fast. In addition to implementing health protocols, other solutions are needed to overcome the high level of spread, namely by making vaccines. The vaccine will stimulate antibodies, thus, the body will recognize the virus and reduce the risk of exposure [3]. Over time, several variants of the coronavirus that cause the disease COVID-19 emerged. On November 26, 2021, based on recommendations from the Technical Advisory Group on Virus Evolution, WHO established a new variant called Omicron [4]. Due to the outbreak of positive cases of the Omicron variant corona, the provision of booster vaccines has intensified. The booster vaccine is the third dose of vaccine given in an effort to break the chain of transmission of COVID-19 with the aim of increasing the body's immunity against the coronavirus [5]. A booster vaccination is considered controversial by the public, so many people give their opinions.

Nowadays, developments in all aspects of life are becoming completely digital. All information is very easy and fast to obtain and disseminate using digital technology. Many people express their opinions through social media. One of the social media that is quite popular for expressing opinions is Twitter. Twitter's active users will reach 206 million worldwide in 2021. While in Indonesia, there are 17.5 million users on the platform [6].

Twitter is a social networking service that allows users to send and read text-based messages, commonly called 'tweets'. Twitter was founded in 2006 by Jack Dorsey. Social media Twitter is widely used by several users to express their opinions, one of them being related to booster vaccines. In this study, data that has been tweeted by Twitter users related to the COVID-19 Booster Vaccine was collected using the Twitter API (Application Programming Interface). Twitter API is useful for easy access to information on the Twitter web.

Referring to the booster vaccine which is considered controversial, in this study, a system will be built to find out whether public opinion on the COVID-19 booster vaccine contains positive or negative sentiments. The result of the sentiment distribution will later be used to determine the performance of the Ensemble Bagging methodData on the distribution of positive and negative sentiments that are too far apart can affect the quality of the data so that it can also affect the performance of the method. In 2003 Nasukawa.T.& Yi conducted a study on sentiment analysis. According to him, sentiment analysis in Indonesian is a technique or method used to identify how sentiment is expressed using text and how that sentiment can be categorized as positive sentiment or negative sentiment. The results of the prototype system achieve high precision (75-95% depending on the data) in finding sentiments on web pages and news articles [7].

Research [8] shows that tweets data classification uses the Naïve Bayes classifier algorithm and support vector machine with the highest accuracy results obtained when using the support vector machine method with an accuracy value of 90.47%. In research [9], sentiment analysis was carried out using the Naive Bayes classifier with an accuracy of 93%. Similar research [10] used the Support Vector Machine and K-Nearest Neighbor methods, although the results were not satisfactory, the use of the SVM method had a better accuracy of 75%. Research [11] conducted research on vaccine actions using the keywords "Sinovac Vaccines" and "Red and white vaccines" using the Naïve Bayes method

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 554−561
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1973

and Support Vector Machine, the positive sentiment results became the sentiment that dominates the two keywods, with the positive precentage results using the Naïve Bayes method for the keyword "vaksinsinovac" is 66% and negative is 34%, while for the keyword "vaksinmerahputih" the percentage of positive sentiment is 89% and 11% for negative sentiment. This study also shows that the average accuracy using the Naïve Bayes method is higher at 85.59%, while the Support Vector Machine method is 84.41%. There is also research [12] related to booster vaccines from the perspective of Indonesian citizens on Twitter using the Naive Bayes method with the keyword "booster vaccine" the results of positive sentiment is 23%, neutral sentiment is 15%, and negative sentiment is 76%, with the average results average accuracy of 89%.

In this study, an analysis of public sentiment towards the COVID-19 Booster Vaccine will be carried out using the ensemble bagging method. Ensemble Bagging is used because it is a combination of a set of machine learning algorithms. The final decision is based on voting from the classification results of the basic models. The basic model used is Naive Bayes, K-Nearest Neighbor, and Decision Tree. The results achieved in this study are related to the results of sentiment analysis and the performance of the Ensemble Bagging method.

2. SEARCH METHODOLOGY

2.1 Research Stages

The steps at this stage start from crawling the data to getting the required dataset. After this step, the data is divided into 2, namely data train and data test. Feature extraction is execute in the data train, then validation is implemented on the basic model classification using cross-validation, then tested on the data test by a meta classifier to make predictions and evaluations. After all these steps, the results of sentiment analysis will be obtained. The following is a flowchart of the design of a sentiment analysis system using the ensemble bagging method on social media Twitter:

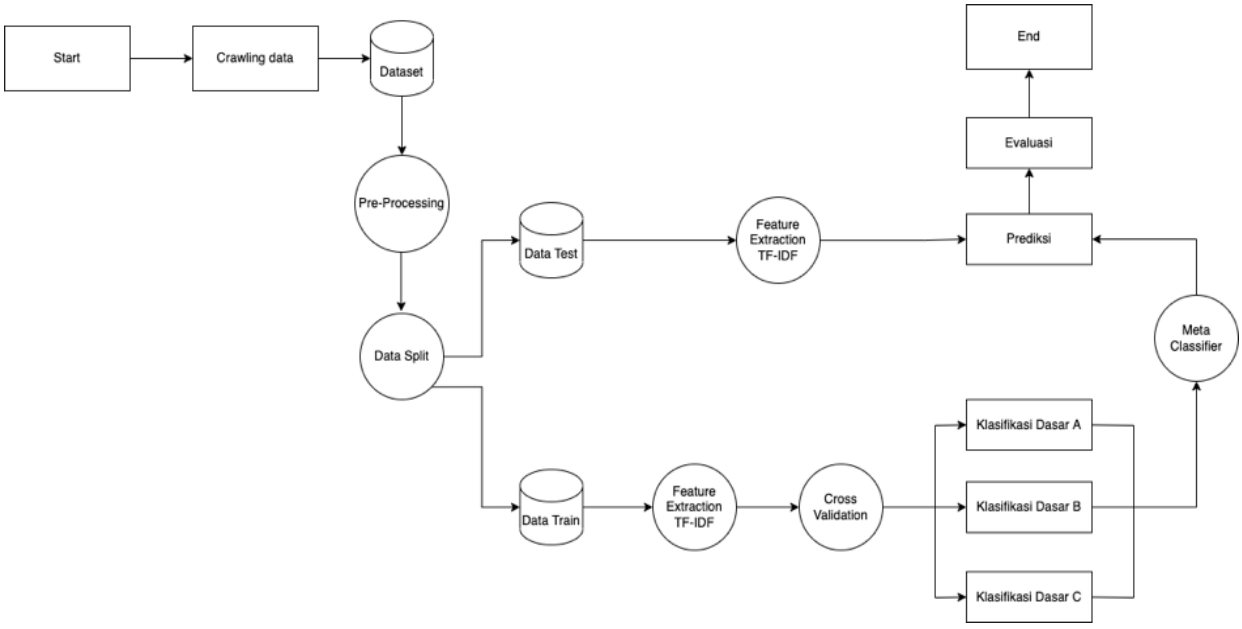

Figure 1. System Model

2.2 Crawling Data

Crawling data is one way to collect data or information. Crawling data can be done in 2 ways, namely search and real-time. In this research, data crawling is done in real-time and requires assistance from the API. The data collection period is from March 2022 to May 2022 and 8752 data are obtained. The information taken is by entering keywords to seek a public opinion regarding the COVID-19 booster vaccine. This information will be used for sentiment analysis which will produce the dataset in this study. After the dataset is collected, a selection is made to obtain data relevant to this study and 6500 selected data are obtained.

2.3 Preprocessing

Preprocessing is carried out to clean the data, one of which is to remove noise, missing values , and inconsistent data. The preprocessing stage is carried out when all the data has been collected. The preprocessing process used in this study, namely Data Cleansing, Tokenization, Stopword Removal, Word Stemming, and Lemmatization.

Cleansing data such as remove mention and hashtag, remove newline, remove emoji, case folding, remove number, remove punctuation and remove white space is used to remove missing values. The newly collected dataset

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 554−561
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1973

must have a missing value or noise. This can happen because the data collection is not perfect, so there are many parts that are missing and irrelevant.
Tokenization is used to divide text in the form of sentences or paragraphs into certain parts.
Stopword Removal is the process of eliminating words that are unimportant or have no meaning and are less influential in the classification process.

Word Stemming and lemmatization is a process of finding basic words by removing affixes that include prefixes, suffixes, or a combination of both, by running a certain algorithm.

2.4 Data Split

Data splitting is a method of dividing the dataset into data train and data test. Data train is data used to train the algorithm, while data test is data used to determine the approximate reference for the performance of the algorithm that has been trained.

2.5 Feature Extraction

Extraction is the process of taking features of an object that can describe the characteristics of the object [13]. The feature extraction used in this research is TF-IDF (Term Frequency-Inverse Document Frequency). In this method, the value of TF and IDF will be calculated for each token using the formula [14]:

$$\boldsymbol{W_{dt} = tf_{dt} * IDF_t} \quad (1)$$

Equation (1) shows W_{dt} is the weight of the d-document against the t word of the keyword, tf_{dt} is the number of words searched for in the d-document against the t word of the keyword, and IDF is $\log\left(\frac{D}{1+d_t}\right) + 1$. The IDF value is obtained from D, the total document, and dt is the number of documents containing the word you are looking for.

2.6 Classification

In this study, we will use the ensemble bagging method and set three basic lessons on the base classifier: Naive Bayes, K-Nearest Neighbor and Decision Tree. The assumption of using the base classifier refers to previous research related to sentiment analysis and getting high accuracy, easy to understand and implement. In the meta classifier stage, input data is trained from the results of the previous base classifier. Predictions will be made from the results of the three previous basic models, and the prediction return will ultimately depend on the three predictions from the base classifier.

2.6.1 Ensemble Bagging

Ensemble learning is a machine learning paradigm where multiple models are trained to solve the same problem and combine to get better results. According to how basic learning is made [15], there are 2 paradigms of the ensemble method, namely the sequential ensemble method and the parallel ensemble method. In ensemble learning, there is an ensemble bagging method. The ensemble bagging method is a method that can improve the results of machine learning classification algorithms by combining prediction classifications from several models.

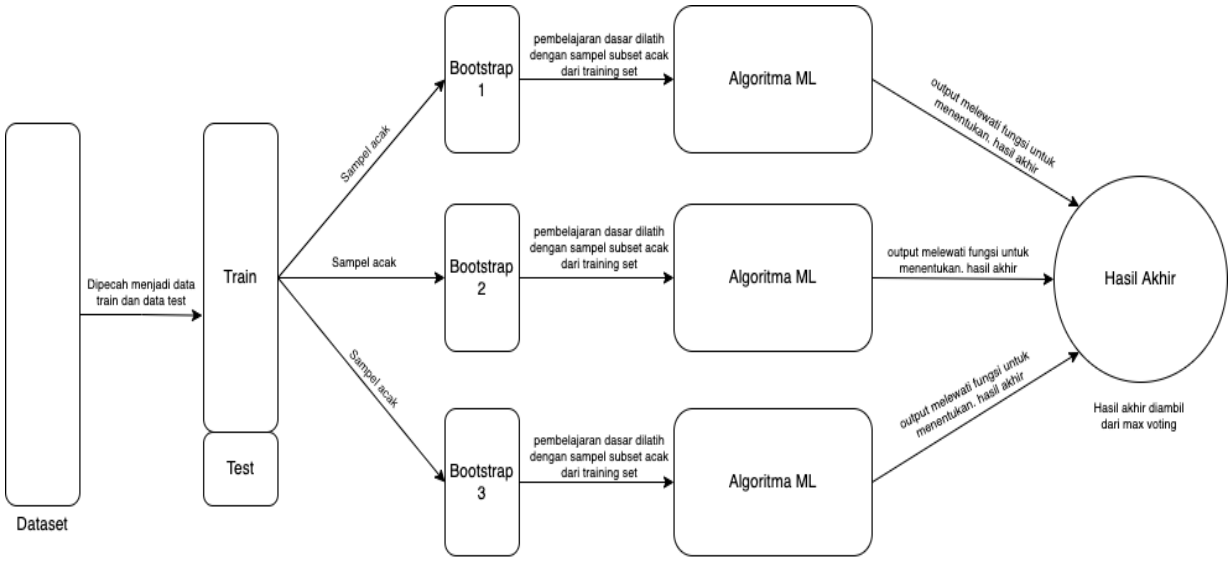

Figure 2. Ensemble Bagging Flowchart

Bagging is an ensemble learning technique that aims to reduce error learning through the implementation of a set of homogeneous machine learning algorithms. Bagging uses a base learner who is trained separately with a random sample of the train data through a voting or averaging approach, resulting in a more stable and accurate model.

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 554−561
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1973

The first bagging process is the extraction of bootstrap subset "n" from the training set, and then the bootstrap subset is used to train base learner "n" with different types. To make predictions, each "n" learner is given a test sample, the output of each learner is taken from max voting.

2.6.2 Naïve Bayes

Naïve Bayes Classifier is the simplest Bayesian learning method. This method produces a statistical classifier based on probability and is a statistical approach to performing inductive inference on classification problems. The following is the calculation formula for Naive Bayes.

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \quad (2)$$

Equation (2) shows P(X|Y) is the probability of the hypothesis based on the conditions, P(Y|X) is the probability based on the conditions of the hypothesis, P(X) is the probability of the hypothesis, and P(Y) is the probability of X.

2.6.3 K-Nearest Neighbor

K-Nearest Neighbor (KNN) is a method for classifying objects based on the learning data that is closest to the object. Learning data is projected into a multidimensional space, and the features of the data are represented from each of these dimensions. The dimension space is divided into several parts based on the classification of learning data. The best value of k depends on the data. In general, a high value of k will reduce the effect of noise on the classification but will make the boundaries between classifications blurry. By using parameter optimization, namely cross-validation, a good value of k can be chosen [16]. Euclidean Distance can calculate the classification results from the closest training data with the following formula:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (y_i - x_i)^2} \quad (3)$$

Equation (3) shows x, y is a point in the Euclidean space, then a Euclidean vector at the origin of the space and n is the number of spaces.

2.6.4 Decision Tree

According to Research [16] the Tree algorithm is commonly used for statistical pattern recognition. In the decision tree, there are three nodes, namely the leaf, the root node (the starting point of the decision tree), and the intermediate node related to the test. A decision tree is a process of repeated calculations until the calculation process can no longer be carried out and all tree attributes have classes [17]. There are several types of decision-makers. Therefore, there are differences between the numerical models (entropy and Gini index) used in the decision tree [18].
Entropy is a measurement of the uncertainty associated with a random variable. The rise and fall of uncertainty or random variables will be directly proportional to the increase in entropy. The value of entropy (E) ranges from 0 to 1 and is calculated as:

$$E = \sum_{i=1}^{c} - p_i \, log_2 p_i \quad (4)$$

Equation (4) shows that p is a non-zero probability that the arbitrary rule belongs to class n. The log function of base 2 is used because the entropy ranges from 0 to 1.

According to research [19] the Gini algorithm is defined for each queue in the information index. Gini evaluates impurities, data partitions, or a set of training levels:

$$Gini = 1 - \sum p^2 \quad (5)$$

Equation (5) shows that p is the value of class i separated by the number of items.

2.7 Evaluation

At this stage, it will be known how the public sentiment on Twitter regarding the COVID-19 Booster Vaccine will be known. The evaluation stage is also used to see the suitability of the model with the data and is measured using a confusion matrix in tabular form. The following is a confusion matrix table:

Table 1. Confusion Matrix

Data		Actual	
		True	False
Prediction	True	TP	FP
	False	FN	TN

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 554−561
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1973

Where:

a. TP is the result of a positive system prediction and corresponds to a positive actual.
b. TN is the result of a negative system prediction and corresponds to a negative actual.
c. FP is the result of positive system predictions but actual negative results.
d. FN is the result of a negative system prediction but the actual result is positive.

It can be observed with a table containing recall, accuracy, precision and F1-Score as a benchmark for the results of the system being built. The following is the recall, accuracy, precision and F1-Score formula used:

a. Recall
Recall is the level of success in retrieving a piece of information by the system.

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

b. Accuracy
Accuracy is the level of closeness between the predicted value and the actual value.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

c. Precision
Precision is the level of accuracy between the information requested by the user and the answer given by the system.

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

d. F1-Score
F1-Score is an evaluation calculation obtained from a combination of recall and precision values.

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

3. RESULTS AND DISCUSSION

3.1 Dataset

The keywords "vaccine booster", "vaccine", and "booster" are used in Indonesian-language tweets. The data crawling process was carried out using the help of the Twitter API in the period from March 2022 to May 2022 and obtained 8752 data. After the dataset was collected, a selection was carried out to obtain data relevant to this research and obtained data from the selection of 6500 with the distribution of positive and negative classes can be seen in Figure 3.

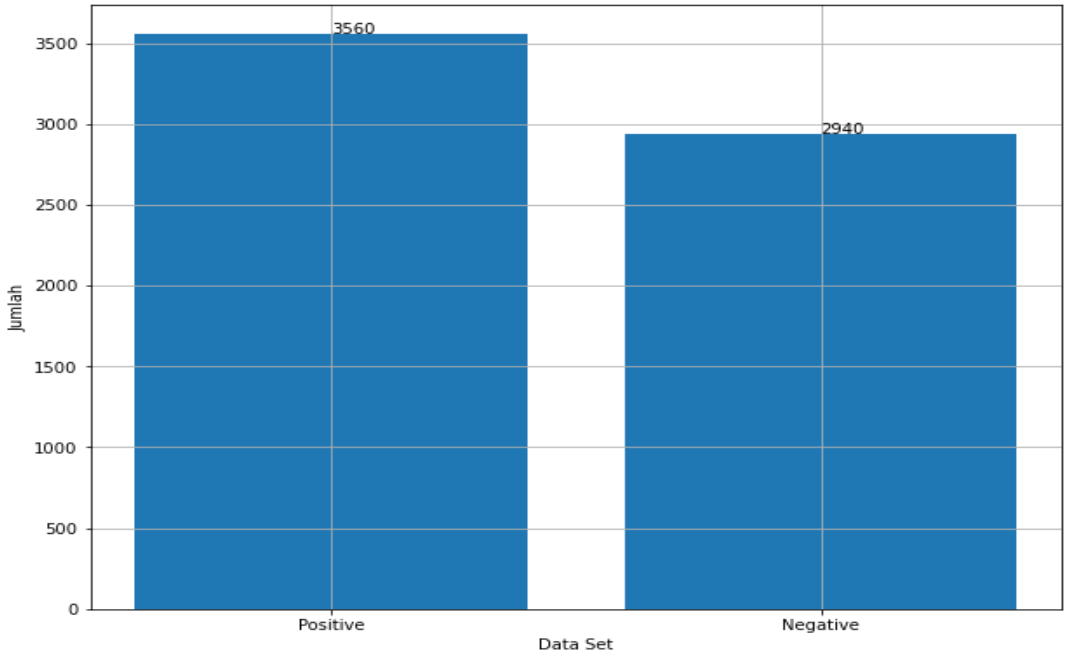

Figure 3. Positive and negative class distribution

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 554−561
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1973

The labeling process was carried out manually by three different people to determine whether a tweet was included in the positive class, which contained praise, satisfaction, and an invitation to do booster vaccines. While the negative class contains insults, innuendo, and feelings of disappointment. Labeling uses the number 1 for positive classes and 0 for negative classes. The results of labeling are exemplified in table 2.

Table 2. Labeling results

Tweet	Label
Saya menolak vaksin lanjutan. Mau booster atau apa. \n\nSudah normal sekarang biarkan ekonomi muter. \n\nKalau mau jualan obat ke China India saja, warganya banyak disana	0
@ArieMutyara Ada kasus ga udah vaksin 1 dan 2 \nTapi ga booster \n\nTiba tiba ada sertifikat vaksin boost di apk peduli lindungi	0
Untung udah vaksin booster dari lama \xf0\x9f\x98\x8c\xf0\x9f\x91\x8d\xf0\x9f\x8f\xb	1
Dia panutan banget sih paling taat prokes, vaksin udah lengkap ditambah booster juga. Semenjak disana kasus cvd naik bgt gak pernah lepas masker pas di yst \xf0\x9f\xa5\xba	1

3.2 Pre-processing

To clarify features, a data cleansing process is carried out during preprocessing, which aims to remove unnecessary attributes, such as mentions, hashtags, new lines, emoji, case folding, numbers, and white space. The results of data cleansing are shown in table 3.

Table 3. Cleansing Data

Before Cleansing Data	After Cleansing Data
Buat yang mau mudik siapkan diri kalian, baiknya lengkapi dosis vaksin sampai booster dan tetap menjaga protokol kesehatan. #AyoVaksinCovid19	buat yang mau mudik siapkan diri kalian baiknya lengkapi dosis vaksin sampai booster dan tetap menjaga protokol kesehatan

After the data cleansing process, tokenization is performed to divide the text into several parts. After that, a stopword removal process is carried out to remove words that are less influential in the classification process. The results of the tokenization and stopword removal processes can be seen in table 4.

Table 4. Tokenization and Stopword Removal

Before Tokenization & Stopword Removal	After Tokenization	After Stopword Removal
buat yang mau mudik siapkan diri kalian baiknya lengkapi dosis vaksin sampai booster dan tetap menjaga protokol kesehatan	[‘buat’, ‘yang’, ‘mau’, ‘mudik’, ‘siapkan’, ‘diri’, ‘kalian’, ‘baiknya’, ‘lengkapi’, ‘dosis’, ‘vaksin’, ‘sampai’, ‘booster’, ‘dan’, ‘tetap’, ‘menjaga’, ‘protokol’, ‘kesehatan’]	‘mudik’, ‘siapkan’, ‘baiknya’, ‘lengkapi’, ‘dosis’, ‘vaksin’, ‘booster’, ‘menjaga’, ‘protokol’, ‘kesehatan’]

3.3 Feature Extraction

After performing the pre-processing stage, feature extraction using the TF-IDF algorithm is used. The data is divided into several ratios, namely 70:30 and 80:20 with n-gram unigram, bigram, and trigram either on stemming data, lemmatization, or not using both.

3.3.1 Comparison of Stemming dan Lemmatization

Table 4 shows the comparison of stemming and lemmatization tweet data and without a stem. At a ratio of 70:30 the best result is a dataset with stemming, while at a ratio of 80:20 the best result is a dataset with lemmatization.

Table 5. Stemming and Lemmatization

Stem Type	Tweet Without Stem	Stem Results
Stemming	mudik siapkan baiknya lengkapi dosis vaksin booster menjaga protokol kesehatan	mudik siap baik lengkap dosis vaksin booster jaga protokol sehat
Lemmatization		mudik siapkan baiknya lengkapi dosis vaksin booster jaga protokol sehat

3.4 Modeling

The research was conducted using three base classifiers in each bootstrap. Based on the validation results, the best dataset results were obtained from the 70:30 and 80:20 train data, which can be seen in Figure 4.

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 554−561
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1973

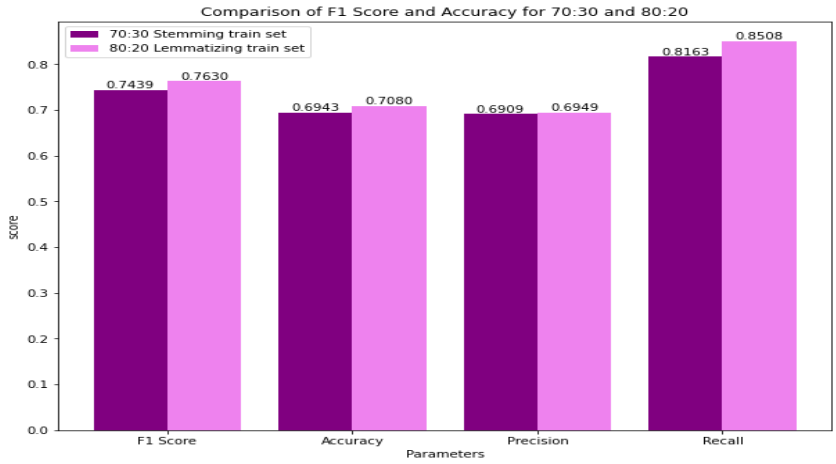

Figure 4. Dataset 80:20

It can be concluded that the dataset with the highest score is the train data with a ratio of 80:20 that has been lemmatized. Then the best dataset is compared again, with a dataset ratio of 80:20 using either unigram, bigram, or trigram.

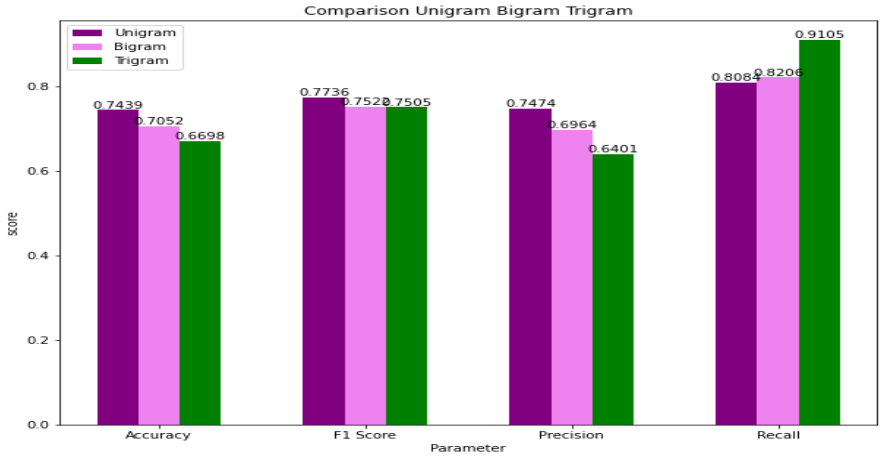

Figure 5. 80:20 Unigram Bigram Trigram

After validation using unigram, bigram, and trigram datasets, it can be seen in Figure 5 that unigram is superior, both in accuracy parameters, F1 Score, and Precision.

3.5 Hyperparameter tuning

In performing Hyperparameter tuning, several parameters are tested on each base classifier using gridsearchCV with a combination of parameters that can be seen in table 6.

Table 6. Hyperparameter Tuning

Base classifier	Parameter grid
Naïve Bayes	var_smoothing = [1.e+00, 1.e-01, 1.e-02, 1.e-03, 1.e-04, 1.e-05, 1.e-06, 1.e-07, 1.e-08, 1.e-09]
KNN	leaf_size = [1, 2, 3, 4, 5, 6]
	n_neighbors = [1, 2, 3, ... , 30]
	p = [1,2]
Decision tree	max_depth = [2, 3, 5, 10, 20]
	min_samples_leaf = [5, 10, 20, 50, 100]
	criterion = [‘gini’, ‘entropy’]

Comparison of the use of the model with hyperparameter tuning and not using hyperparameter tuning and the results of the model with hyperparameter tuning add to the score F1 score, accuracy, and precision. However, the use of hyperparameter tuning slightly reduces the recall score of the whole model.

3.6 Performance

In the evaluation of the model, the ensemble bagging method was used and three basic lessons were determined in the basic classification, namely Naive Bayes, K-Nearest Neighbor, and Decision Tree. In the base classifier, the

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 554−561
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1973

highest accuracy value is using Naïve Bayes of 0.72. As for the ensemble, the highest score was obtained using hyperparameter tuning, which can be seen in table 7.

Table 7. Performansi Metode

Method	Precision	recall	F1 Score	Akurasi
Naïve Bayes	0,72	0,73	0,72	0,72
KNN	0,66	0,65	0,65	0,67
Decision Tree	0,67	0,66	0,66	0,67
Ensemble dengan Hyperparameter Tuning	0,72	0,71	0,72	0,72
Ensemble tanpa Hyperparameter Tuning	0,72	0,70	0,70	0,72

4. CONCLUSION

This study applies the Ensemble Bagging method in conducting sentiment analysis. The data tested were 6500 with the keywords "vaccine booster", "vaccine", and "booster" in Indonesian-language tweets. In the period from March 2022 to May 2022, 3560 positive tweets and 2940 negative tweets were obtained. The performance results on Naïve Bayes are slightly higher than on Ensemble with Hyperparameter Tuning, with the recall value on Naïve Bayes is 0.73 and Ensemble is 0.71. This is because the other two base classifiers, namely KNN and Decision Tree, do not perform very well, so they can affect the max voting in the Ensemble Bagging aggregation process.

REFERENCES

[1] "Pertanyaan dan jawaban terkait Coronavirus," World Health Organization, 2022. https://www.who.int/indonesia/news/novel-coronavirus/qa/qa-for-public (accessed Mar. 30, 2022).
[2] "Coronavirus Disease 2019 (COVID-19) Situation Report – 87," World Health Organization, Feb. 16, 2022. https://cdn.who.int/media/docs/default-source/searo/indonesia/covid19/external-situation-report-87_16-february-2022.pdf?sfvrsn=a5169f8b_5 (accessed Mar. 30, 2022).
[3] "4 Manfaat Vaksin Covid-19 yang Wajib Diketahui," UPK Kemenkes, 2021. https://upk.kemkes.go.id/new/4-manfaat-vaksin-covid-19-yang-wajib-diketahui# (accessed Apr. 04, 2022).
[4] "Informasi Terbaru tentang Omicron," World Health Organization, Nov. 30, 2021. https://www.who.int/indonesia/news/detail/30-11-2021-informasi-terbaru-tentang-omicron (accessed Apr. 04, 2022).
[5] dr. Rizal Fadli, "4 Manfaat Vaksin Booster yang Perlu Diketahui," halodoc, Feb. 08, 2022. Manfaat Vaksin Booster yang Perlu Diketahui (accessed Apr. 04, 2022).
[6] "Leading countries based on number of Twitter users as of January 2022," Statista Research Department, Mar. 22, 2022. https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/ (accessed Nov. 22, 2021).
[7] "AF IDA, KHUSNA," 2018. http://eprints.umpo.ac.id/4324/3/BAB%20II.pdf (accessed Dec. 03, 2021).
[8] F. Fitriana, E. Utami, and H. al Fatta, "Analisis Sentimen Opini Terhadap Vaksin Covid-19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes," Jurnal Komtika (Komputasi dan Informatika), vol. 5, no. 1, May 2021.
[9] W. Yulita, E. D. Nugroho, and M. H. Algifari, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier," Jurnal Data Mining dan Sistem Informasi, vol. 2, no. 2, pp. 1–9, 2021.
[10] A. Baita, Y. Pristyanto, and N. Cahyono, "ANALISIS SENTIMEN MENGENAI VAKSIN SINOVAC MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN K-NEAREST NEIGHBOR (KNN)," Information System Journal (INFOS), vol. 4, no. 2, 2021.
[11] B. Laurensz and E. Sediyono, "Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19," Jurnal Nasional Teknik Elektro dan Teknologi Informasi, vol. 10, no. 2, 2021, doi: https://doi.org/10.22146/jnteti.v10i2.1421.
[12] H. Khoiril, F. Abd. Charis, and Harliana, "Sentiment Analysis of Vaccine Booster during Covid-19: Indonesian Netizen Perspective Based on Twitter Dataset," JTKSI, vol. 5, no. 2, 2022.
[13] D. Satria and Mushthofa, "Perbandingan Metode Ekstraksi Ciri Histogram dan PCA untuk Mendeteksi Stoma pada Citra Penampang Daun Freycinetia," Jurnal Ilmu Komputer Agri-Informatika, vol. 2, no. 1, pp. 20–28, 2013.
[14] A. A. Maarif, "PENERAPAN ALGORITMA TF-IDF UNTUK PENCARIAN KARYA ILMIAH," 2015.
[15] Z.-H. Zhou, Ensemble Methods Foundations and Algorithms. Taylor & Francis Group, 2012.
[16] F. Liantoni and H. Nugroho, "Klasifikasi Daun Herbal Menggunakan Metode Naive Bayes Classifier dan K-Nearest Neighbor," Jurnal SimanteC, vol. 5, no. 1, Dec. 2015.
[17] A. Novantriani, M. K. Sabariah, and V. Effendy, "Analisis Sentimen pada Twitter untuk mengenai Transportasi umum darat Dalam Kota dengan Metode Support Vector Machine," e-Proceeding of Engineering, vol. 2, no. 1, p. 1177, Apr. 2015.
[18] P. A. Octaviani, Y. Wilandari, and D. Ispriyanti, "Penerapan Metode Klasifikasi Support Vector Machine (SVM) pada Data Akreditasi Sekolah Dasar (SD) di Kabupaten Magelang," Jurnal Gaussian, vol. 3, no. 4, pp. 811–820, Oct. 2014, doi: https://doi.org/10.14710/j.gauss.v3i4.8092.
[19] O. Rahmati, M. Avand, P. Yariyan, J. P. Tiefenbacher, A. Azareh, and D. T. Bui, "Assessment of Gini-, entropy- and ratio-based classification trees for groundwater potential modelling and prediction," Geocarto International, pp. 1–20, Feb. 2021, doi: 10.1080/10106049.2020.1861664.